

Question 1:

At the end of phase I, in terms of deliverables, are you expecting a software application OR a written method (in Word/pdf etc.) to accomplish the program goals? Could you please be a little more specific about the deliverables after Phase I (1 year), Phase II (2 years) and III (2 years)?

Answer 1:

As stated in the BAA, the Program requires delivery of a prototype system (that is, a "software application") at the 12-month point of Year 1. This system should be sufficiently robust to demo the capabilities explored and to make clear the contribution and direction of the research performed.

The deliverables for the two option years being offered in the current BAA are dependent on the research topic proposed. These deliverables should reflect advancements in concepts, cultures, groups, data types addressed and significant improvements and expansion of the systems developed. It will be especially important to show willingness to work with other teams in the Program in order to develop integrated resources for the intelligence analysts.

The deliverables for the two years following the first 3 years will be specified in a BAA that will be issued at the end of the 3 years.

Question 2:

My understanding is that Section 2.2.4.3 is intended to be an example of "the social science theories that help to define the social features" and I do not find any reference to social science theories. Also I do not find a clear differentiation between sections 2.2.4.1. and 2.2.4.3.

Answer 2:

Section 2.2.4.3 provides examples of topics that social scientists have explored. Section 2.2.4.1 identifies social features or goals of individuals that are members of groups. These latter features can, therefore, be considered manifestations of the larger social frameworks of interest to social scientists.

Question 3:

Since "social science" is an extremely broad term, more examples of specific theories would be very helpful. For example, do the terms of this announcement strongly discourage Ethnographic approaches which generally do not focus on universal characterization? Also since the announcement cites Levinson and Brown, can I assume that Linguistic approaches are sanctioned?

Answer 3:

The goal of the SCIL Program is to gain from the insights, systems, theories, and constructs that the social sciences and social scientists (e.g., psychologists, anthropologists, sociologists, sociolinguists) have already developed. What social

science, social theory or social constructs are selected depends on the topic chosen for research. The notion of social theory should not be viewed too rigidly.

Question 4:

I hate to be repetitive, but perhaps my greatest confusion is in understanding how "automated" fits in to this announcement. Most of the linguistic features identified assume socio-cultural analysis has already taken place, and if it has taken place, then where does automation fit in? Perhaps, your intention was something along these lines:

- a.) Use linguistic features defined independently of their socio-cultural role (e.g., phrasal categories, word classes, pronouns, thematic relations).
- b.) Use some sort of machine learning algorithm or rule-based system to discover correlations between linguistic features and some predefined set of socio-cultural construct/feature (such as "deference").
- c.) Now that you have "socio-cultural+linguistic" features (such as "honorific verbal inflection"), these can be further leveraged to discover correlations with more complex/composed "socio-cultural construct/features" (such as "group cohesion")?

Answer 4:

Many approaches to the research are possible. The fundamental idea is to show how language correlates with and manifests the social dimensions of groups, such as the formation and development of a group, the relationships in it, the (in)stability of the group, the roles and status of the members of a group of people. How this is understood and implemented is for the researcher to propose.

Question 5:

"Non-standard language" (Section 2.2.4.2): Do you mean low prestige varieties such as AAVE, Creoles, or Pidgins? Or do you mean minority languages (which would include the overwhelming majority of the world's languages)? Or do you mean things like colloquial speech?

Answer 5:

"Non-standard language" is meant as a general term and includes any form of a language that is not of high register. The important concept here, however, is how variants of language are used to accomplish particular social goals, whether standard or non-standard.

Question 6:

Are FFRDC's eligible for submitting proposals for the subject BAA?

Answer 6:

FFRDCs may take part in the SCIL Program, as long as they follow the guidelines for FFRDC participation outlined in the FAR (Federal Acquisition Regulation). For example, FFRDCs may only participate as a sub-contractor on a team.

Question 7:

Could you clarify what is meant by "Social goals?" Do they relate to social functioning within a group (such as "politeness")? Or could they refer to the plans of a group (such as "blowing up a nuclear plant")?

Answer 7:

A *social goal* is generally a tacit social behavior. In the SCIL Program, a social goal is meant to be one that is evidenced by the use of particular language. *Planning* as a manifestation of a concept such as *intent* would be included but identification of the steps that lead to a particular event would probably not address the goals of the Program. It would be up to the proposer to show how such a topic would have general applicability.

Question 8:

Linguistic 'features' such as honorifics, greetings (and I assume leave-taking), register, politeness markers, are defined in terms of their socio-cultural function (i.e., the thing we are trying to connect the linguistic 'feature' to). Hence the detection of socio-cultural function based on these linguistic features seems redundant. If a text has feature 'honorific', I already know its socio-cultural function.

Should I assume that this circularity is inadvertent or that these 'features' were included to encourage investigators to explore their automatic detection rather than the automatic calculation of their signification?

Answer 8:

It is not clear that the assumptions of this question hold. An *honorific*, for example, refers to a class of lexical items that can represent varying degrees of respect. To label a text as *honorific* would therefore be insufficient to capture the use of the term. Detecting terms without specifying the social contribution of the terms is not of interest.

Question 9:

Can I assume the following potential roles for automated techniques are allowable under the terms of this program:

- a) Automatic Identification of socio-culturally significant linguistic features.

In example 2 of BAA Appendix I, this would amount to automatically detecting that the choice of 'usted' over 'te' is correlated with husband and wife (i.e., a social structure). This would just be detecting that linguistic feature is significant in some way.

- b) Automatic association of linguistic features to specific socio-cultural significance.

This would amount to automatically detecting that the choice of 'usted' over 'te' signals "intimacy". This would involve both detecting that a linguistic feature is significant and understanding its significance.

- c) Automatic calculation of significance.

Given that I know that a linguistic feature is significant, I calculate its significance by either mapping the feature onto a fixed set of possible significations or by generating hypotheses about possible significations and choosing one of them. I am essentially referring to the second half of role #2 (i.e., understanding significance).

d.) How important is the commitment to a particular theory of Social Science and conversely, to what extent can the results (i.e. output of the computational system) be pre-theoretic (i.e., agnostic with respect to the particular constructs of a theory)?

Answer 9:

Although all options presented are possible, Option C would be a far more interesting approach. The use of *Usted* and *tu* in the example given is important because the norm is to use *tu* as a marker of intimacy. That norm is violated, however. If it is not clear why it is violated, a faulty interpretation is likely (e.g., that the pair having marital problems or is angry with one another).

The goal of the SCIL Program is to ensure that the assumptions about language use are informed by the valuable work done in the social sciences and not to recreate that work. The adherence to a particular theory and its constructs may not prove completely viable when the features are automated. It is the responsibility of the proposer, however, to make clear how any work done has been influenced by/informed by principles of a relevant social science.

Question 10:

Do we need a partner to respond?

Answer 10:

It is the responsibility of the proposer to make clear how the proposed research activities will be informed by concepts from the social sciences. It is assumed that such information will be best obtained by working with a social scientist with relevant experience on the research topic of interest. If the responder has such experience, that must be made clear.

It is also assumed that a social scientist would require the services of a computer scientist so that the social concepts can be shaped appropriately for automation.

Question 11:

I'd like to know if there are scenarios available for the problem domain that would help put the proposed R&D into perspective?

Answer 11:

It is believed that the BAA offers enough information about the topics of interest and the diversity of approach allowed to be sufficient.

Question 12:

We presume the interest is primarily in transcripts of spoken language, is that correct?

Answer 12:

Spoken language is of interest but other types of interactions are also permitted, including e-mail, blogs, chat, instant messaging, speeches. Each of the data types introduces significant advantages and disadvantages. It is incumbent on the proposer to make clear how the data selected will support the research question, what the advantages and disadvantages of the data type are and how the anticipated results might tackle other data types, other languages, etc.

Question 13:

Is there an existing data set that will be used for this program?

Answer 13:

No. See Section 2.3 in the BAA. It is the responsibility of the proposer to select the data and argue for its relevance to the particular research proposed.

Question 14:

At the core of this program is the fact that the linguistic structure of body linguistic events is sensitive to the social context of the events. It would seem that the parsing algorithms might benefit strongly from previously given information about that context, as well as providing new information in a way suited to inclusion with previous information. For instance, utterances that indicate possession of objects might in one context indicate high status and in another indicate low status for the same object. Thus there would seem to be a potentially significant positive role for a module that negotiates this context with the parsing module? Would a proposal that incorporates such a module be of interest?

Answer 14:

Clearly, the larger physical context and situation of an interaction can contribute to the understanding of the exchange. The SCIL Program is focused on the use of language, however. If it can be demonstrated that a "context module" contributes substantially and critically to the advancement of the program goals and to the automation of language use, it could be considered.

Question 15:

I have not seen an official PIP, will one be issued? Will the full proposal have a standard format (e.g., 25 pp. tech proposal, plus independent cost volume, etc.?)

Answer 15:

See the following web sites for further details:

http://www.nbc.gov/acquisition/fort_h/solicit.html

<http://www.fbo.gov/spg/DOI/OS/SAB/08%2DSCIL/listing.html>

Question 16:

Is there any pre-proposal White paper?

Answer 16:

No

Question 17:

I would like to inquire whether Israel would be eligible to apply for this grant and participate in this program. In addition, can an Israeli researcher apply to be placed on the bidders list for the SCIL Program?

Answer 17:

All are welcome to submit proposals for participation in the SCIL program.

Question 18:

Can partners from the EU be included?

Answer 18:

All are welcome to submit proposals for participation in the SCIL program.

Question 19:

We are wondering if it is possible to arrange a meeting with the PM for discussion about SCIL?

Answer 19:

No, direct contact with individuals responsible for the program is not permitted prior to official selections.